DOCUMENT RESUME

ED 414 752 FL 024 947

AUTHOR Pearson, Jennifer
TITLE Electronic Texts and Concordances in the Translation Classroom.
PUB DATE 1996-00-00
NOTE 12p.; For serial publication in which this article appears, see FL 024 940.
PUB TYPE Journal Articles (080) -- Reports - Descriptive (141)
JOURNAL CIT TEANGA: The Irish Yearbook of Applied Linguistics; n16 p85-95 1996
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; Classroom Techniques; *Electronic Text; Foreign Countries; Graduate Study; Higher Education; *Indexes; Information Sources; Language Research; *Languages for Special Purposes; *Reference Materials; Second Language Instruction; Second Languages; *Translation
IDENTIFIERS Dublin City University (Ireland); *Language Corpora

ABSTRACT

Ways in which language corpora and concordancing tools can be used in classes in translation or languages for special purposes are examined. Use of these tools is recommended as a useful or even essential complement to a conventional dictionary-based approach, especially considering the increasing availability of large volumes of text in electronic format and reasonably priced tools for text analysis. The first section of the article presents a profile of graduate-level translation students at Dublin City University (Ireland) and outlines the level of expertise they are expected to achieve. Normal translation practice and some common problems encountered by students in using conventional dictionaries are outlined. Subsequently, the types of electronic resources now available to those students are described, and ways they can be exploited successfully are explained. This section also includes a description of available software tools. Focus is on retrieval of information about a term, its meaning, and its usage. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 414 752

Electronic texts and concordances in the translation classroom

Jennifer Pearson
Dublin City University

This article examines some of the ways in which corpora and concordancing tools can be used in the context of LSP teaching, with particular emphasis on the specialised translation syllabus. It aims to demonstrate that a corpus-based approach to translation is a useful and perhaps even an essential complement to the more conventional dictionary-based approach. The increasing availability of large volumes of text in electronic format and reasonably priced tools for exploring and analysing texts may prove to be invaluable in the teaching of translation. I have already used some of these resources in the translation classroom and found that students are i) extremely receptive to the notion of applying modern technology to translation practice and ii) much less likely to make incorrect terminological choices when they source their material in electronic text.

The first section of this article contains a profile of translation students at Dublin City University and outlines the level of expertise that they are expected to have. Normal translation practice is described, as are some of the problems encountered by students when trying to locate the meaning of a word or phrase in one language and its equivalent in another through consultation of dictionaries. The following section describes the types of electronic resources that are now available to students at Dublin City University and it suggests a number of ways in which they can be usefully exploited. This section also includes a description of the software tools that are available. The final section contains examples of how these resources and tools can be exploited in order to retrieve information about a term, its meaning and its usage.

FL021947

BEST COPY AVAILABLE

Introduction

The students under discussion are final year students on the Applied Languages Programme, and postgraduate students on the MA in Translation Studies Programme at Dublin City University. These students

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Jeffrey Kallen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

are required to translate, into their mother tongue, specialised texts dealing with aspects of what we broadly term economics, and science and technology. While first and second year undergraduate students are given 3 hours a week of lectures, in English, on economics and science and technology as background to the specialised subjects that they encounter in the translation classroom, final year and masters students are expected to source their own background material. Frequently, these students have little prior understanding of the subject matter and they are encouraged to read up on the subject before attempting a translation. They have a number of ways of doing this. If translating a text on natural gas heating systems, for example, they may call into their local gas company office to speak to one of their experts. They may know someone who is working in the area who can explain the basic concepts to them. They may find what they need in the university library. In addition to sourcing background documentation, students are encouraged to locate comparable texts in their mother-tongue. If, for example, they are translating product specifications for a cooker from French into English, they should look for similar specifications in English.

When students have reached an understanding of the subject matter of a text, they then start the translation process. Their first step is to identify equivalent target language terms for terms in the source language text that are unfamiliar to them. They generally consult a bilingual general or specialised dictionary for this purpose. They are advised to double check on the appropriateness of the equivalents retrieved by consulting monolingual dictionaries in the source and target languages. However, even when students follow the recommended procedures, a number of problems arise with the consultation of dictionaries. The dictionaries which they are likely to consult can be broadly subdivided into the following- categories: monolingual general language dictionaries, monolingual specialised dictionaries, bi- or multilingual general language dictionaries and bi-or multilingual specialised dictionaries.

Monolingual general language dictionaries will provide a definition of the entry, multi-word variants of the entry. They may also contain some phraseological information, indicating some of the typical collocates for the entry. However, monolingual general language dictionaries are of limited use in specialised translation as they will rarely contain definitions of specialised terms; where definitions of technical terms are provided, they tend not to contain sufficient detail. As with monolingual general language dictionaries, monolingual specialised dictionaries provide definitions; they are likely to indicate multi-word variants of the term and they may indicate associated terms. However, they do not provide

phraseological information, leaving students to guess the correct collocate and all too often, they get it wrong. The layout of entries in general and specialised bi- or multilingual dictionaries tends to be fairly similar. Equivalents are provided for the entry itself and for multi-word variants of the entry. In the larger dictionaries, phraseological information will also be provided. However, no definition is provided, with the result that students frequently mistranslate, selecting the incorrect reading when more than one is provided.

While we can draw our students' attention to these issues and alert them to potential problems which at least makes them more wary when consulting dictionaries, there are many instances where dictionaries will simply not contain the information which students are seeking. Take acronyms (i.e. abbreviations such as GATT, SSA, SIPTU) for example which are particularly common in economics texts. If the acronym has only been coined recently, it will not appear in a dictionary and it is possible that native speakers may be unable to help. The only option remaining is to locate comparable texts in the hope of retrieving the full form of the acronym. This is time consuming and may not even yield the desired result. Take for example, the French acronym TIPP, which this author first encountered in a newspaper article on budget proposals in France. It did not appear in the Dictionary of Acronyms. French native speakers were unable to help. The text itself provided no clue as to the meaning of the word other than that it was likely to be the acronym for a recently introduced tax or levy. Consultation of the Le Monde CD-ROM provided the information. When the levy was first introduced, the full form usually appeared in brackets after the acronym. TIPP stood for a tax on imports of petroleum products (Taxe sur l'Importation des Produits Pétroliers). Another area where dictionaries may be found to be wanting is in the definition of words associated with a new or emerging technology. The meaning of the word may have evolved since publication of the dictionary. Thus, a student seeking a definition of CD-ROM may find that it is defined as being used for the storage of written text whereas it is in fact now used for the storage of sound, film, graphics as well as for the storage of text. Alternatively, the word may simply not appear in the dictionary at all. Take the English term *dongle*, a software protection device that plugs into the back of a computer; this term is not to be found in any of the large range of specialised dictionaries available to students at DCU.

Electronic resources and concordances

This brief sketch of the problems associated with relying exclusively on the conventional dictionary-based approach to translation highlights how important it is to encourage students to look further afield in their search

for the meaning of a term or the appropriate equivalent in another language.

In recent years, DCU has acquired a number of CD-ROMs which are proving to be of enormous benefit to students. These resources serve as a one-stop shop for background information on projects for language and other subjects. Students are encouraged to view these electronic repositories as potential terminological resources and to observe the Firthian principle that *you shall know a word by the company it keeps*. In essence, this means that if students examine the way in which a word is used and have sufficient instances of the word in actual usage, an analysis of these instances should enable them to understand the meaning of the word, locate related words and identify appropriate collocates.

We will now look at a number of different methods of eliciting information about a word from a corpus or from a single text. Baker (1995:2) says of a corpus that:

> A corpus now means primarily a collection of texts held in machine readable form and capable of being analysed automatically or semi-automatically in a variety of ways.

The two resources used for this article are Encarta, the Microsoft Encyclopaedia on CD-ROM and the 1992-1995 editions of the Financial Times available on CD-ROM. These are both collections of naturally occurring running text, i.e. the language has not been tampered with or edited.

The software used to retrieve the information is the MicroConcord concordance software developed by Tim Johns and Michael Scott at the University of Birmingham. A concordancer is a piece of software which allows users to retrieve all occurrences of a particular word or phase, known as the *node*, in a corpus together with the segment of text in which the node is located. This segment is called the concordance line and is generally 80 characters in length, i.e. the width of an A4 page, but can be extended if users wish to retrieve information that lies beyond the scope of the 80 characters.

Table 1 Concordance of *coal*

are a new generation of advanced	coal	utilization processes, some of
higher heating value. Anthracite	coal	has the highest carbon content
lowest rank of coal. Bituminous	coal	has even more carbon and a corr
ennsylvania. The best bituminous	coal	for coking purposes comes from
ects on the environment. Burning	coal	produces carbon dioxide, among
new clean coal technology (CCT).	Coal	liquefaction supplies all of th
South Africa's oil needs. Clean	Coal	Technologies CCTs are a new ge
ed in gasification and new clean	coal	technology (CCT). Coal liquefac
as improved methods of cleaning	coal,	fluidized bed combustion, inte
urization. Location of Deposits	Coal	is found in nearly every region
chemical by-products, including	coal	tar, which are used in the manu
teel producers use metallurgical	coal,	or coke, a distilled fuel that
dioxide (SO2) emissions from new	coal-fired	facilities have been cont
pure carbon. Other components of	coal	are volatile hydrocarbons, sulf
producing coal. Various types of	coal	are classified according to fix
at, due to the widespread use of	coal	and other fossil fuels, the amo
first stage in the formation of	coal,	has a low fixed carbon content
r in lignite, the lowest rank of	coal.	Bituminous coal has even more
metric tons. Of this recoverable	coal,	China held about 43 percent, t
its of lignite and subbituminous	coal	in North Dakota, South Dakota,
rals that remain as ash when the	coal	is burned. Some products of coa
Also, sulfur and nitrogen in the	coal	form oxides during combustion t
oped until the 20th century. The	coal	reserves of the United States a
issions have dropped even though	coal	use has increased. All ranks of
arbon content and heating value.	Coal	may be transformed by further p
antic states. Estimates of world	coal	reserves vary widely. According
	Coal,	solid fuel of plant origin. In

Table 1 contains an example of the concordance for *coal* extracted from the Microsoft Encyclopaedia Encarta. Once a concordance has been produced it can be sorted in a number of ways, i.e. to the left or right of the node; the concordance for *coal* is sorted alphabetically one word to the left of the node. In a large concordance of several hundred lines, users can select only those patterns which are of interest to them. For example in a concordance of the word *take*, the user may only be interested in occurrences of *take* when by the preposition *of* appears two words to the right of the node. This allows the user to retrieve *take care of, take account of, take notice of* but to exclude *take place, take the dog for a walk, take a driving test.*

A concordancer will also allow users to list all variants of a word, in which case they simply specify the lemma + wildcard. For example, they could specify 'comput*' if they wished to retrieve words such as *compute(s,d), computing, computer(s), computation, computational* etc.

Using concordancers

The next section will look at how concordancing software can be used to retrieve different types of terminological information. The first approach involves identifying the meaning or scope of a term by using the text as the sole resource, i.e. without recourse to a dictionary. This is particularly

BEST COPY AVAILABLE

useful in situations where students simply do not have access to the dictionaries which they would normally require for this purpose.

Students are asked to examine a text manually and to identify all occurrences of a particular term within the text. They are then asked to make a note of any phrases, clauses or sentences that appear to describe the meaning of the term. On the basis of these instances alone, it is often possible for students to write a definition of a term, to identify related terms and locate phraseological information about the term. Students are then asked to compare their result with the definition of the term in specialised dictionaries. They are invariably surprised at how much more complete their own definition is, particularly in relation to related terms and phraseological information about the term itself. Once they have carried out the task manually, they are then shown how the same information can be retrieved much more easily by using a concordancer.

For the purposes of illustration we have chosen to look at the term *coal* as this is a term with which readers will be familiar and should therefore allow them to judge the adequacy of the results obtained. However, the approach is in fact most useful in situations where students are trying to retrieve information about terms which are unknown to them. The concordance for the term *coal* (Table 1) reveals some very interesting information. For example, types of coal are clearly identifiable. If the nouns which appear to the left of the node are combined with the node, the combined result is frequently a subordinate or hyponym of the node. Thus, there are references to *bituminous coal*, *subbituminous coal*, *anthracite coal*, among others. This is a simple means of establishing genus-species relations. It is easy to identify these using a concordancer but a manual analysis of the text for retrieval of the same information would take a lot more time and students might simply overlook some of the references.

When the node is immediately followed by a noun or nouns, the noun(s) which follow(s) is/are the head of a multi-word term relating to coal. Thus, we read of *coal liquefaction techniques*, *coal production*, *coal combustion*, *coal utilization processes*. This gives the reader information about *processes* relating to coal. Furthermore, there are many markers which indicate that the text may contain some definitional information about the node itself. For example, we find: *coal: solid fuel of plant origin*. What follows in the original text, to which the student can refer by expanding the concordance to a full sentence or paragraph, is actually a definition of how coal is formed. In another concordance line we find a reference to *coal and other fossil fuels* so we know that *coal* is a type of fossil fuel.

An analysis of the verbs which co-occur with *coal* will tell the students what types of process *coal* is likely to undergo. Thus, we see that it is *produced, burned, transformed, cleaned, consumed,* all of which is invaluable information to the student looking for the correct collocate. When all of the information which has been retrieved is collated, it can be stored on a terminological record sheet (cf Table 2) for future reference.

Table 2 Terminological record sheet

Term: coal

Grammatical information: Noun, generally in singular form.

Hypernym: fossil fuel.

Definition: Solid fuel of plant origin, found all over the world. Commercial deposits confined to Europe, Asia, Austrial, North America. Peat is first stage in formation of coal. Coal classified according to fixed carbon content. Coal burning produces carbon dioxide, among other by-products. Components of coal include volatile hydrocarbon, sulfur and nitrogen.

Hyponyms: bituminous coal, anthracite coal, metallurgical coal, coke, subbituminous coal, recoverable coal.

Related terms: coal utilization processes, coal-using processes, clean coal technology, coal liquefaction techniques, coal liquefaction supplies, coal deposits, coal production, coal combustion, coal reserves, coal use.

Co-occurs with : formation of, produces, is found, may be transformed, burning, combustion, consumed.

The next time students are asked to translate a text on coal production, they can consult their record sheet to identify appropriate collocates. Coal is *produced*, it is not manufactured. Coal is *formed*, it does not evolve. Coal may be *transformed*, but not converted. The collocates may seem obvious but it is surprising how frequently students allow the language used in a source text to interfere and lead them to use the incorrect collocate in the TL text. The record sheets which students use are generally bi- or trilingual and include equivalents in each of the source languages, information relating to gender, collocates and related terms in the source languages.

The second approach is one which has already been mentioned and entails using a corpus in the same way as one uses a dictionary. There is no need for a concordancer in this instance. Students simply access the corpus and use a key word search to locate the word or phrase they are looking for. This is particularly useful for locating the full form of an acronym, such as TIPP (Taxe sur l'Importation des Produits Pétroliers) which was mentioned earlier, or CCT (clean coal technologies) which appears in the text on coal (Table 1). When students use the corpus for this purpose, they are treating the corpus in the same way as a dictionary; they are simply looking for the full form of an acronym, in order to make a decision about how to treat it in translation.

The third approach involves using the corpus as a means of establishing whether there is a gap between the dictionary definition of a word and the way in which it is used in text; this is an exercise which belongs more in the context of economic or general translation than in that of scientific or technical translation where meaning shifts are less frequent. We have chosen to look at instances of the word *sleaze* (Table 3) to establish whether there has been in a shift or narrowing in the meaning of this word in recent years and also to demonstrate that it is not always appropriate to consult a dictionary particularly when one is dealing with contemporary texts. When students were asked to suggest words that they associated with *sleaze,* they suggested words such as prurient, dirty, sordid, slippery, slimy which was not surprising as these associations were borne out by the readings in the Oxford English Dictionary of *sleaze,* namely:

sleaze (sli:\), *sb.slang.* Back-formation from SLEAZY, SLEEZY a.]
1. Squalor; sordidness, sleaziness, dilapidation; (something of) inferior quality or low moral standards. Also *attrib.*

> **1967** *Listener* 14 Sept.326/2 For all its brazen sleaze, Soho is a pretty fair working model of what a city neighbourhood should be. **1975** *Publishers' Weekly* 29 Dec 68/2 Obviously written to cash in on 'Mandingo', this isn't even readable sleaze: the plot's sloppy, Gilchrist hasn't the knack for writing commercial sex, and the hero is too despicable to be seductive. **1976** *National Observer* (U.S.) 17 July 16 (*heading*) At home with the sleaze king. **1981** *New Yorker* 9 Mar. 104/1 These stores are vast, computerized sleaze centers, where you can buy almost anything - pills, toys, candy, liquor, stockings, pillows and gadgetry.

2. A person of low moral standards.

> **1976** *Telegraph* (Brisbane) 3 Aug 10/3 When I made the mistake of calling them 'sleazy' to their faces, their reactionrwas outrage. 'Don't call me a sleaze,' said Miss Currie. **1977** *Time* 28 Feb 48/1 Oh God, red nail polish - I look like a sleaze.

The examples cited are revealing. *Sleaze* is directly or indirectly associated with Soho, commercial sex and red nail polish!

Table 3 Concordance of *sleaze*

says there has been no sign of post-	**sleaze**	puritanism. 'Not a sausa
amid continuing allegations of	**sleaze**	against ministers, Mr Major wa
ster wishes to calm concerns of	**sleaze**	hanging over his government.
h the continuing allegations of	**sleaze**	against Conservative MPs by an
affair - a long-running case of	**sleaze**	far more serious than any yet
night to defuse the charges of	**sleaze**	levelled against his governmen
is investigating allegations of	**sleaze**	against Tory MPs, last night v
ion to shake off the charges of	**sleaze**	that have damaged his governme
anished. However, a new form of	**sleaze**	has emerged in the 1980s and 1
Britain has a notable record of	**sleaze**	A century of hypocrisy in high
g pressure over accusations of	**'sleaze'**	the MPs - who make up the Lab
y way to address the growth of	**'sleaze'**	in government. Amid continuin
vestigate fresh allegations of	**'sleaze'**	after reports connecting two
with rebutting allegations of	**'sleaze'**,	as the national Conservative
government over allegations of	**'sleaze'**	has raised concern about the
to recover from allegations of	**'sleaze'**	against senior Tory MPs, Lord
polls, enmeshed in charges of	**'sleaze'**	and unable to demonstrate gri
ations in the British press of	**'sleaze'**	in public life have put a las
of the specific allegations of	**'sleaze'**	that have been made against m
---------- A stench of	**'sleaze,**	cronyism, insider dealing, wa
oming tainted by allegations of	**sleaze.**	Speaking at prime minister's
d sinking in a deep blue sea of	**sleaze.'**	A former government official
ident in a British tradition of	**sleaze.**	Another concern is the curren

Regular reading of the English newspapers a couple of years ago revealed that *sleaze* was generally being used in rather a different way than the above examples might suggest. Table 3 above contains a random selection of citations of the word *sleaze* from the 1992-1994 editions of the Financial Times, preceded by the preposition *of.*

What does the concordance reveal? There are allegations of sleaze against ministers, charges of sleaze levelled against the government; Britain has a notable record of sleaze. These are scarcely the types of collocates which one would have expected on the basis of the dictionary definition of sleaze. *Sleaze* co-occurs with *cronyism, insider dealing.* There is a tradition of *sleaze* in Britain. Government Ministers, Tory MP's, a former government official are all associated with it. Elsewhere in the concordance, we also find that *In Britain, there is sleaze, not corruption.* This suggests that *sleaze* is the equivalent of *corruption* when it applies to the British government. It appears to be something that involves politicians and members of the Tory government in particular. There are no references to *sex,* to *red nail polish,* to areas like *Soho.* The only physical place with which the word is associated is the Parliament. Further examination of the *concordance* reveals that the word is actually loosely defined:

BEST COPY AVAILABLE

Three broad categories of illicit practice are evident. The first could be called 'sleaze' - dubious practice on the margins of impropriety, involving relatively small sums and favours.

This confirms our intuitions. *Sleaze* is specifically a *dubious practice on the margins of impropriety, involving relatively small sums and favours.* In the period 1992-1994, all occurrences of *sleaze* in the FT referred to *sleaze* as corruption, with references to cash-for-questions, to members of the government. It is clear from only a cursory examination of the concordances that *sleaze* is being used specifically to refer to corrupt practices. Interestingly, in the 1991 edition of the FT, the word *sleaze* was still being used exclusively to refer to something sordid. There is no reference to allegations or charges of sleaze. It is only in early 1992 that the new meaning starts to emerge. The reason why this particular example is included here is that it clearly demonstrates that, for the translation of texts dealing with current affairs, students would do well to consult parallel contemporary sources when these are available rather than to rely exclusively on dictionaries which may not be sufficiently recent to have taken account of a shift in meaning.

Conclusion

This article set out to demonstrate that the availability of huge volumes of text in electronic form and the development of concordance software for manipulating information which is retrieved from these collections of text can enhance the work which students have to undertake when preparing to translate a text. It highlighted some of the pitfalls of the dictionary approach and showed how corpora can be used as a means of identifying terminological information. The intention was to show that a text can tell as much about a term as any specialised dictionary and even more besides if one considers that a text provides phraseological as well as definitional information. Furthermore, corpora are valuable uptodate repositories which contain information which might never find its way into a dictionary. Concordancing software allows users to retrieve several instances of a word or phrase in order to draw conclusions about its meaning.

Bibliography

Baker, M. (1995). Corpora in Translation Studies. *Target*, Volume 7 (forthcoming).

Financial Times L.d. (1992-1994). *The Financial Times on CD-ROM*, FT Profile, London.

Microsoft Corporation (1995). *Encarta*. United States, Microsoft.

Simpson, J.A., Weiner, E.S.C. (1989). *Oxford English Dictionary*. Clarendon Press, Oxford.